Response Surface Split-Plot Designs;
A Literature Review

Luis A. Cortes, The MITRE Corporation
James R. Simpson, JK Analytics
Peter A. Parker, NASA Langley Research Center

Abstract

The fundamental principles of experiment design are factorization, replication, randomization, and local control of error. In many industrial experiments, however, departure from these principles is commonplace. Often in our experiments, complete randomization is not feasible because factor level settings are hard, impractical, or inconvenient to change, or the resources available to execute under homogeneous conditions are limited. These restrictions in randomization result in split-plot experiments. Also, we are often interested in fitting second-order models, which lead to second-order split-plot experiments.

Although response surface methodology has experienced a phenomenal growth since its inception, second-order split-plot design has received only modest attention relative to other topics during the same period. Many graduate textbooks either ignore or only provide a relatively basic treatise of this subject. The peer-reviewed literature on second-order split-plot designs, especially with blocking, is scarce, limited in examples, and often provides limited or too general guidelines. This deficit of information leaves practitioners ill-prepared to face the many challenges associated with these types of designs. This article seeks to provide an overview of recent literature on response surface split-plot designs to help practitioners in dealing with these types of designs.

1

1. Introduction

Fisher[1] laid down the fundamental principles of experiment design: factorization, randomization, replication, and local control of error. Factorization consists of making deliberate, simultaneous changes to the experimental factors to find the individual or mutual effect those factors have on the response variables of interest. Replication is the application of a combination of experimental factors, called treatments, to the experimental units to obtain a valid estimate of the experimental error. Randomization refers to the random assignment of treatments to experimental units and to the random assignment of experimental runs to treatments. Randomization averages out the effects of undesirable factors present in the experiment, generally enables the assumption that the experimental errors are independent and identically distributed random variables, and produces an unbiased estimate of variance. Local control of error is an experiment design technique that allows for minimizing the influence of nuisance factors on the response by partitioning the experimental units into homogeneous subsets called blocks. Local control improves the precision of the comparison of factors of interest and reduces or eliminates the component of the variability transmitted from nuisance factors.

Sometimes it is impractical or impossible to carry out an experiment while adhering simultaneously to all the principles of experiment design. For instance, many industrial experiments may involve situations in which the complete randomization of experimental runs may not be feasible because of factor level settings that are impractical or inconvenient to change, limitations in the resources available to complete the experiment in homogeneous settings, or both. There are experiment that involve combinations of materials that are rare-to-find as well as easy-to-find. Other experiments involve the application of some treatments to large experimental units as well as the application of other treatments to smaller experimental units. Similarly, some experiments require the batch processing of experimental units for some factors but not for other factors, or the stream processing of different experimental units requiring the application of the same treatment. Likewise, some experiments involve factors that need to be estimated more precisely than other factors, or a combination of factor levels that are hard-to-change as well as easy-to-change. Practitioners generically refer to these situations as experiments with hard-to-change factors and easy-to-change factors, and to the experiments themselves as *split-plot*

experiments. Experiments that require fitting higher-order models under restricted randomization conditions are referred to as *response surface split-plot experiments*.

This article seeks to provide an overview of recent literature related to response surface split-plot designs. It starts with a retrospective quick-look at the seminal work on split-plot experiments by Fisher[2] and Yates[3,4] and on the defining feature of the split-plot designs. Then the article highlights the classical work on response surface methodology by Box and Wilson[5] and its application to split-plot design. Reviewed in more detail are key contributions to response surface split-plot design by Letsinger, Myers, and Lentner[6], Bisgaard[7], Goos[8], Vining, Kowalski, and Montgomery[9], Parker, Kowalski, and Vining[10,11,12], Jones and Nachtsheim[13], Vining[14], and Myers, Montgomery, and Anderson-Cook[15].

2. Split-Plot Designs

Split-plot experiments originated in agronomic research. In agronomic experiments, some factors like irrigation method are restricted to large areas of land, called whole-plots. Whole-plots are split into smaller areas of land, called sub-plots, which allow for the individual application of treatments, such as seed variation (some experimenters refer to sub-plots as split-plots). Factors associated with whole-plots are called whole-plot factors while factors associated with sub-plots are called sub-plot factors. Because the whole-plots are split into sub-plots, there is more experimental material for the whole-plots than for the sub-plots. Fisher[2] provided the first example of "analysis of variation" in experimental field trials to analyze the effect different fertilizers had on the yield of potato.

Like in agronomic experiments, in industrial split-plot experiments the hard-to-change factors are associated with the whole-plots while the easy-to-change factors are associated with the sub-plots. There are two different randomizations in split-plot experiments. First, the combination of hard-to-change factors, or whole-plot treatments, are randomly assigned to the whole-plots. Then, within each whole-plot, the combination of easy-to-change factors, or sub-plot treatments, are randomly assigned to the sub-plots. Consequently, a split-plot experiment can be thought of as a superposition, or nesting, of two experiments—one experiment based on the whole-plot treatments and another experiment based on the sub-plot treatments.

Blocking is both a technique for controlling error and a form of restricted randomization. In the context of local control, blocking improves the precision of the comparison between factors by arranging the treatments into groups, or blocks, that have similar sources of variability that are extraneous to the experiment. The differences in variability between the blocks due to irrelevant sources are identified and removed analytically leaving in the experimental error only the differences within treatments in the same block. In the context of restricted randomization, the whole-plots are analogous to blocks except that there is interest in understanding the variability between whole-plots. There are many variations of the split-plot design such as the split-split-plot design, the split-block design, the strip-block design, and the split-block split-block design. Goos[8], Vining[14], Hinkelmann and Kempthorne[16], Federer and King[17], Wu and Hamada[18], and Montgomery[19] discussed split-plot design and its variations in detail. Kowalski, Parker, and Vining[20] provided a tutorial on industrial split-plot experiments.

Due to the restriction in randomization, the distinguishing feature of a split-plot designs is a model with two error terms, which in matrix form is:

$$y = X\beta + \delta + \varepsilon \qquad (1)$$

where y represents the vector of responses, $\mathbf{X}$ represents the model matrix that includes both the whole-plot and the sub-plot terms, β represents the vector of regression coefficients, δ represents the vector of whole-plot error terms, ε represents the vector of the sub-plot error terms, and $\delta + \varepsilon \sim N(0, \Sigma)$. The error terms incorporate measurement error, variability from uncontrolled factors, variability in the experimental units to which the treatments are applied to, general background noise, etc. Figure 1 illustrates the error structures of a 2^3 completely randomized design and a 2^3 split-plot design (with whole-plot factors A and B and sub-plot factor C).

Bisgaard and de Pinho[21] showed that the whole-plot factors and their interactions have a larger variance than the sub-plot factors and their interactions. The variance of the whole-plot factors and their interactions has two components—one component coming from the whole-plot and another coming from the sub-plot. The variance of the sub-plot factors and their interactions with other factors has only one component, which is coming from the sub-plot. In general, for k

factors and any 2^k factorial with N runs, p whole-plot factors, $q = k - p$ sub-plot factors, the variance for the whole-plot factors and their interactions is

$$\sigma^2_{wp-factors} = \frac{4}{N}\left(2^q \sigma^2_\delta + \sigma^2\right) \qquad (2)$$

Similarly, the variance for the sub-plot factors and any interaction with them is:

$$\sigma^2_{sp-factor} = \frac{4}{N}\sigma^2 \qquad (3)$$

In general, due to the randomization restriction, the estimates involving sub-plot factors are more precise. Split-plot experiments provide less information on the whole-plot factors relative to the same factors in similarly sized completely randomized experiments. However, a gain on information on the sub-plots effects and the sub-plot by whole-plot interactions compensates for the loss of information on the whole-plot. The terms $\sigma^2_{wp-factors}$ and $\sigma^2_{sp-factors}$ are called variance components, and are components of the total variability in the observations. Montgomery[19] provided a basic discussion on variance components while Searle, Casella, and McCulloch[22] provided a more complete discussion.

Despite of the utility and advantages offered by split-plot experiments, practitioners may confront many of the roadblocks that surround experimentation under restricted randomization:
- the structure of split-plot experiments is significantly more complex than the structure of comparable completely randomized designs;
- sometimes the experiment run matrix is not executed as planned and results in a split-plot design, which adds complexity to the analysis;
- sometimes split-plot experiments are not recognized as such and are inadvertently analyzed as if they were carried out as completely randomized experiments, which results in inaccurate models;
- the selection of the correct error term for testing the significance of the factors and their interactions is sometimes unclear, especially when the whole-plot and sub-plot factor effects are aliased;

- some split-plot designs are undesirably large, even for a reasonably small number of factors;

- sometimes the analysis of variance method produces a negative estimate of the interaction component, which results in a sub-plot variance component larger than the whole-plot variance component;

- whole-plot replication is necessary for estimating the whole-plot error term, which may increase the cost of the split-plot experiment relative to a completely randomized experiment;

- a typical approach for reducing the cost of split-plot experiments is to place more factors at the whole-plot level, which reduces the power to detect significant effects.

Fortunately, the peer reviewed literature addressing the roadblocks mentioned above is extensive and dates, as already mentioned, to the cradle of experiment design methodologies. Practitioners can find in the literature multiple examples that can help them address those issues.

Yates[3] discussed the confounding of main effects and orthogonality in split-plot experiments. Since each whole-plot contains only one whole-plot treatment, the differences between treatments coincides with the difference between whole-plots. Hence, they are confounded. Thus, there is only trivial information for the comparison between whole-plots while the comparison within whole-plots is unaffected by the confounding. In experiments with unreplicated whole-plots, only the error term associated with the sub-plots can be estimated and it cannot be used to test for significance of the whole-plot factors. Thus, the additional error components along with valid estimation of each error component are key concepts in split-plot experiments. Yates[4] discussed the structure and analysis of split-plot experiments: (1) the random assignment of blocks to whole-plots; (2) the random assignment of whole-plots to sub-plots; (3) the random assignment of sub-plots to observational units; and (4) the random selection of observational units.

Huang, Chen, and Voelkel[23] constructed minimum aberration split-plot fractional factorial designs, which minimize the number of main effects aliased with low-order interactions. Bingham and Sitter[24] used the minimum aberration criteria to rank some designs created by combining a fractional factorial design at the whole-plot level with a fractional factorial design at the sub-plot

level while Bingham and Sitter[25] provided theoretical results. Bingham and Sitter[26] exemplified the effect of restricted randomization on the choice of split-plot design for industrial applications while Loeppky and Sitter[27] discussed the analysis of those experiments. Because minimum aberration designs have a large number of whole-plots with a small number of sub-plots, they are viewed unfavorably for use in industrial applications.

Addelman[28] used split-plot confounding to construct 2^k x 2^q factorial and 2^{k-p} x 2^{q-r} fractional factorial split-plot experimental plans. Bisgaard[7] provided a comprehensive tutorial on fractional factorials in a split-plot structure using the aliasing structure as criteria for selecting a design instead of minimum aberration. Using split-plot confounding to take advantage of using different sub-plot designs was a significant contribution and an important step forward for using fractional factorials split-plots for industrial applications.

To reduce the number of experimental runs in split-plot designs, Kulahci and Bisgaard[29] provided split-plot design construction techniques using two-level Plackett-Burman[30] designs, which are particularly helpful for screening experiments. Tyssedal and Kulahci[31] simplified the analysis of designs by Kulahci and Bisgaard[29] and showed that their analysis can be done using ordinary least squares (OLS) regression. Tyssedal, Kulahci, and Bisgaard[32] constructed two-level split-plot designs where the sub-plots were run as mirror image pairs, which separate the sub-plot main effects and sub-plot by whole-plot interactions from the rest. Kulahci and Tyssedal[33] provided a two-level split-plot design construction methodology for multistage experiments using the Kronecker product representation of 2^k designs, which can be used for any number of stages and different number of sub-plots for each stage.

Hader[34], Wooding[35], Box[36], Simpson, Kowalski, and Landman[37], and Vining[14] all recognized the role of split-plot experiments in industrial experiments and warned us that often they are not recognized as such and are incorrectly analyzed as if they were completely randomized designs. Daniel[38] referred to this misapplication as an "inadvertent split-plot". The consequences of an inadvertent split-plot are the mixing of the whole-plot error and the sub-plot error, which inflates the variance of the regression coefficients in the model for the sub-plot factors and masks the effects of the whole-plot factors, which in turn produces erroneous tests of significance. Based on the specifics of the design, the tests of significance for second-order split-plot designs are tests

for purely sub-plot effects, tests for effectively whole-plot effects, and tests for effects somewhere in between.

Equally important as the inadvertent split-plotting issue is the failure to reset the factor levels between consecutive runs and then analyze the experiment as if resets did occur. Ganju and Lucas[39,40] illustrated how a situation like this produces inappropriate tests of significance. Ju and Lucas[41] demonstrated that split-plot blocking can provide superior sub-plot parameter estimates as compared to completely randomized designs. Webb, Lucas, and Borkowski[42] compared the prediction properties of completely randomized experiments to experiments where the factors levels were not reset. They described how the failure to reset the factor levels in successive runs can result in less precision of the parameters estimates and in inflated prediction variances.

Bisgaard[7] coined the term Cartesian product to describe the scalar product technique that Taguchi[43] used for crossing the inner factors array and the outer factors array in his product array experiments. The technique is like the bi-randomization design construction techniques proposed by Letsinger, Myers, and Lentner[6] and like the robust product design construction technique introduced by Box and Jones[44]. Taguchi's experiments are constructed using a full Cartesian product method, which often results in large, unpractical designs. Also, as pointed out by Bisgaard[7], the Taguchi's product array experiments are not widely recognized as split-plot experiments and are incorrectly analyzed as if they were completely randomized designs, which often results in incorrect models. Bisgaard and Sutherland[45] showed that Taguchi's famous INA Seito tile manufacturing experiment was indeed a split-plot design and reanalyzed it using a standard split-plot approach.

While incorrectly analyzed, the Taguchi INA Seito tile manufacturing experiment showcased the utility of design of experiments. Taguchi identified eight active ingredients in the tile clay and only with 16 runs he found a clay formulation that resulted in a more robust product. That experiment became one of the industrial experiments that motivated Japan to embrace the Taguchi system and propelled the country to become a world-wide leader in product development for years to come. Taguchi's inner and outer array experiments are highly regarded by the quality and manufacturing control community. Most likely because of their success, they stifled the intellectual development of the design and analysis of split-plot experiments. It was not until the

8

turn of the millennium, coincident with Letsinger, Myers, and Lentner[6], Bisgaard[7], and Box and Jones[46], that split-plot experiments began to gain due recognition as a valuable method for industrial experiments.

While the peer reviewed literature is extensive on two-level split-plot designs, the literature for response surface split-plot design is scarce, has limited examples, and provides guidelines that are often too general. The next section provides a literature review on response surface designs as background to the literature review on response surface split-plot designs that follows.

3. Response Surface Designs

Box and Wilson[5] catalyzed the application of response surface methodology to industrial experiments. In many industrial experiments, it is necessary to fit a second-order model to the observations to reveal the true underlying process conditions or product characteristics. In the case of a completely randomized design, all factors have the same importance or the same value and the second-order Taylor series approximation model used to fit the observations takes the form:

$$E(y) = \beta_0 + \sum_{i=1}^{k} \beta_i\, x_i + \sum_{j<i}^{k} \sum_{i=2}^{k} \beta_{ji} x_j x_i + \sum_{i=1}^{k} \beta_{ii}\, x_i^2 \qquad (4)$$

where y represents the response, x_i represents the ith independent variable or factor to which the observational units are subjected to, and the $\beta's$ represents the regression coefficients that are estimated empirically.

The most popular response surface designs are the central composite design (CCD), by Box and Wilson[5], and the Box-Behnken design (BBD), by Box and Behnken[47]. Another popular response surface design is the 3^k general factorial design, which is generally inadequate for many industrial applications because of its size.

3.1. Central Composite Designs

The central composite design (CCD) is the workhorse of response surface methodology. CCDs are five-level designs, which for k factors consist of a combination of 2^k factorial or 2_V^{k-p}

fractional factorial designs, center points, and axial points located a distance α from the center of the design. To fit a full second-order model with k factors, central composite designs require $2^k + 2k + k_0$ design points (k_0 is the number of center points) and can estimate $(k + 2)(k + 1)/2$ coefficients. CCDs have circular ($k = 2$), spherical ($k = 3$), or hyper-spherical ($k > 3$) symmetry. The distance α of the axial points from the center of the design—a function of the number of factors, their levels, and the desired properties of the design—determines the features and properties of the design. A spherical CCD is a design in which all the factorials and axial points are on the surface of a sphere or radius $\alpha = \sqrt{k}$. This means that the prediction variance of the design is the same at all points that are at the same distance from the center of the design. In other words, the variance of the predicted response does not change when the design is rotated about the center point. Designs with this property are named rotatable designs. Rotatability is a property that can be used for the selection of a response surface design. Box and Hunter[48] discussed the rotatability of second-order response surface designs. Box and Hunter[48] also derived orthogonal blocking (block effects do not affect the ability to estimate the model coefficients independently) arrangements for several types of designs, including an arrangement to block a CCD in two orthogonal blocks.

3.2. Box-Behnken Designs

Box-Behnken designs (BBD) are a family of three-level rotatable or near rotatable designs. The construction technique relies on using balanced incomplete block designs and 2^k factorial designs. The design avoids the corners of the design space in favor of edge points located at the mid-level ($x_i = 0$) of the factor levels, which results in poor estimation at the factorial point locations. Thus, BBDs are more useful for situations in which there is no interest in predicting at the factorial points of the cube. Like for CCDs, replicated runs at the center points permit a more uniform estimation of the prediction variance over the design space. Although practitioners often associate the BBD with cuboidal regions because of its cubic appearance, the BBD is a spherical design. Because these designs are rotatable or near-rotatable, they require sufficient center points to improve their prediction accuracy.

3.3. Small or Saturated Response Surface Designs

There are situations in which scarce resources—funds, time, material, manpower, or equipment—makes it impractical to allow the use of standard designs for fitting second-order models, especially when the number of factors k is high. For those situations, small or specialized response surface designs could be attractive. Some of the most popular small or saturated response surface designs were proposed by Hartley[49], Westlake[50], Rechtschaffner[51], Doehlert[52], Hoke[53], Pesotchinsky[54], Lucas[55], Box and Draper[56], Roquemore[57], Mitchell and Bayne[58], Notz[59], Draper[60], Morris[61], Oehlert and Whitcomb[62], and Gilmour[63].

3.4. Optimal Designs

Keifer[64,65] and Keifer and Wolfwitz[66] laid the foundation for evaluating and comparing designs based on optimal design theory—designs that are "best" with respect to some criterion. Optimal designs are a good option whenever it is inadequate to use classical designs.

3.5. Definitive Screening Designs

Jones and Nachtsheim[67] proposed a class of three-level screening designs when the number of factors is $k > 5$. Definitive screening designs (DSD) provide estimates of the main effects that are uncorrelated with two-factor interactions and pure quadratic terms. Because they are three-level designs, the quadratic effects are estimable. DSDs require $2k + 1$ runs. Two-factor interactions are only partially confounded with other two-factor interactions as opposed to Resolution IV screening designs in which two-factor interactions are completely confounded with other two-factor interactions. Pure quadratic effects are not completely confounded with interactions.

Jones and Nachtsheim[67] also provided an algorithm to calculate the pairwise correlation coefficient between two model terms. Jones and Nachtsheim[68] expanded the application of DSD to any number of two-level categorical factors. Jones and Nachtsheim[69] developed flexible orthogonal blocking DSD strategies for situations involving only quantitative factors and for situations involving a mix of quantitative and two-level qualitative factors. For both fixed and random blocks, the numbers of blocks may vary from two to k and the block sizes do not need to be equal. Jones[70] examined the goodness of DSDs.

Errore *et al.*[71,72] showed that DSD can break down when there are several active second-order terms. Jones and Nachtsheim[73] showed that tailoring the model selection to the structure of a DSD, orthogonality of main effects and orthogonality between main effects and second-order effects, can lead to a better model-selection.

4. Response Surface Split-Plot Designs

In a split-plot experiments, the assumption that all factors have the same importance or the same value does not hold, and the second-order Taylor series approximation takes the form:

$$E(y) = \beta_0 + \sum_{i=1}^{p} \beta_i z_i + \sum_{j<i}^{p} \sum_{i=2}^{p} \beta_{ji} z_j z_i + \sum_{i=1}^{p} \rho_{ii} z_i^2 + \sum_{i=1}^{q} \gamma_i x_i + \sum_{j<i}^{q} \sum_{i=2}^{q} \gamma_{ji} x_j x_i + \sum_{j=1}^{p} \sum_{i=1}^{q} \gamma_{ji} z_j x_i + \sum_{i=1}^{q} \theta_{ii} x_i^2 \quad (5)$$

where x_i represents the ith sub-plot factor, z_i represents the ith whole-plot factor, $k = p + q$, β represents the regression coefficients for the whole-plot linear terms, ρ represents the regression coefficients for the whole-plot pure quadratic terms, γ represent the regression coefficients for the sub-plot linear terms and whole-plot by sub-plot interaction terms, and θ represent the regression coefficients for the sub-plot quadratic terms.

Myers[74] provided a review of, and outlined the status of, response surface methodology. Anderson-Cook *et al.*[75] covered response surface design evaluation. Neff and Myers[76] reviewed the impact of recent developments in response surface methodology on applications in industry. Myers *et al.*[77] reviewed the developments in response surface methodology from 1989 through 2004, including split-plot experiments, and synthesized the state-of-the-art and areas for research in robust parameter design, response surface designs, multiple responses, generalized linear models, and other topics. The paper presented a brief historical perspective, identified three extensive reviews conducted over the last 50 years, and provided an extensive bibliography. Khuri and Mukhopadhyay[78] surveyed the development of response surface methodology and provided research directions. Khuri[79] reviewed the application of response surface methodology to the agricultural and food sciences.

4.1. Restricted Central Composite Designs

Lucas and Ju[80] studied completely randomized, completely restricted (two equally sized replicates), and partially restricted (four equally sized replicates) run orders using a CCD with three whole-plot factors and one sub-plot factor. The runs were equally divided into two blocks. They found that for restricted randomization, the residual standard deviation was much smaller and all regression coefficients except the linear and quadratic coefficients for the whole-plot factors have much smaller standard deviations.

Myers, Montgomery, and Anderson-Cook[15] provided an example of a restricted CCD with two whole-plot factors and two sub-plot factors. Parker, Kowalski, and Vining[11] provided a catalog of balanced and unbalanced CCDs. Wang, Vining, and Kowalski[81] studied the rotatability of CCDs in a split-plot structure and proposed a two-strata rotatable split-plot CDD where the prediction variance is a function of the distance to the whole-plot center and the distance to the sub-plot center separately.

English, Simpson, Landman, and Parker[82] characterized the flight performance of a small-scale unmanned aerial vehicle developed for commercial and military operations using a minimum whole-plot CCD provided by Parker, Kowalski, and Vining[12]. The experiment involved one hard-to-change factor (wing tip height) and two easy-to-change factors (angle-of-attack and yaw angle).

4.2. Restricted Box-Behnken Designs

Myers, Montgomery, and Anderson-Cook[15] provided an example of a restricted BBD with two whole-plot factors and two sub-plot factors. Parker, Kowalski, and Vining[11] provided a catalog of balanced and unbalanced BBDs.

4.3. 2^k + Center Runs, 3^k, 4^k, and Subset Designs

There are situations where it is impractical to fit a second-order model with a classical design because it requires an impractical high number of treatments, it has a complex alias structure, or it cannot accommodate constraints associated with block structures, block sizes, or a reduced number of runs. Designs that overcome those disadvantages are needed.

Simpson, Kowalski, and Landman[37] studied the effects that four factors—front ride height, rear ride height, yaw angle, and grille configuration—had on the aerodynamic performance of a NASCAR Winston Cup Chevrolet Monte Carlo stock car in a wind tunnel. The levels for the front ride height and rear ride height factors were hard-to-change while the levels for the yaw angle and grille configuration factors were easy-to-change. The layout consisted of one two-level, replicated whole-plot at each of the four factorial points $(z_1, z_2) = (\pm 1, \pm 1)$ augmented with one whole-plot at the whole-plot center $(z_1, z_2) = (0, 0)$. The sub-plot structure was similar and consisted of one sub-plot run at each of the four factorial points $(x_1, x_2) = (\pm 1, \pm 1)$ and one whole-plot at the whole-plot center $(x_1, x_2) = (0, 0)$. The center points allowed for testing and isolating curvature at both the whole-plot and sub-plot level. Replication provided degrees-of-freedom for estimating both the whole-plot variance and the sub-plot variance. The model contained terms for all linear effects, for all two-factor interactions, for the confounded sub-plot quadratic terms $\beta(x_1^2 + x_2^2)$, and for the confounded whole-plot quadratic terms $\beta(z_1^2 + z_2^2)$. Although the type of situation presented by the authors was a departure from traditional experiment design, it is commonplace in industrial experiments and emphasized the importance of adapting traditional response surface methods to fit specific needs while preserving the desirable properties of response surface designs.

Another approach to construct split-plot designs is stratum-by-stratum. In multi-stratum experiments, sets of experimental units must have the same treatment for at least one factor (like a split-plot design). This strategy can produce designs with more than one strata (i.e. split-split-plot designs) and can be used to fit second-order or higher response surface models. This strategy produces whole-plots that are not balanced and designs that are less efficient than D-optimal designs. Cheng[83] provides a procedure for constructing multi-stratum designs and discusses the properties of those designs.

Gilmour and Trinca[84] provided practical advice on the regression analysis of blocked response surface designs, including an alternative approach to calculate the estimates for variance components as well as the recommendation to use REML for estimating the regression coefficients. Trinca and Gilmour[85] used a second-order split-plot design to maximize the yield and purities of two proteins in a protein-extraction process. They considered one hard-to-change factor (feed position) and four easy-to-change factors (feed flow rate, gas flow rate, and the concentration of two proteins).

Gilmour[63] provided four-level response surface designs based on irregular two-level fractional factorials. Gilmour[86] introduced a class off three-level response surface design called subset designs, which are constructed by using 2^k designs in subsets while holding the other factors at the center. Mylona, Macharia, and Goos[87] introduced two new families of three level GLS-OLS equivalent-estimation (more on that later) split-plot designs—one based on the three-level subset designs introduced by Gilmour[86] and the other based on supplementary difference set designs. Goos and Gilmour[88] discussed a general strategy for construction multi-stratum and split-plot designs. Gilmour and Trinca[89] discussed statistical inference for blocked designs. Trinca and Gilmour[90] provided an algorithm to improve the efficiency of multi-stratum designs while Trinca and Gilmour[91] applied the optimality criteria to obtain multi-stratum designs and showed that these designs have better properties for inference than a comparable optimal design.

4.4. Fractional Factorial Designs

To construct orthogonal fractional factorial split-plot designs with two or more levels, Sartono, Goos, and Schoen[92] used an approach involving integer linear programming and mixed linear programing for small design problems. Similarly, they used integer linear programming and variable neighborhood search for large design problems.

4.5. Optimal Designs

Goos and Vandenbroek[93] developed an algorithm for constructing D-optimal split-plot designs. Goos[8] addressed additional aspects of optimal designs in a blocked and split-plot structure while Goos and Vandenbroek[94] constructed D-optimal split-plot designs with specific numbers of whole-plots.

Jones and Nachtsheim[13] provided a comprehensive review of split-plot designs and proposed a D-optimal split-plot design algorithm that trades replicates at the center points of the whole-plot and sub-plots for sub-plot runs that are at the corners of the design region.

Myers, Montgomery, and Anderson-Cook[15] provided an I-Optimal design for an experiment involving two hard-to-change factors (cure temperature and percent of resin in the adhesive) and two easy-to-change factors (amount of adhesive and cure time) that affect the

strength of an adhesive used in a medical application. Jones and Goos[95] showed that both the prediction capability and the estimation of factor effects of I-Optimal response surface split-plot designs is substantially better than D-Optimal response surface split-plot designs.

4.6. Small or Saturated Response Surface Split-Plot Designs

Parker, Kowalski, and Vining[12] provided a catalog of Box and Draper[56], Hoke[53] near saturated, Notz[59] saturated, and hybrid minimum whole-plot designs.

4.7. Definitive Screening Designs

Lin and Yang[96] studied the performance of DSDs adapted to split-plot experiments for one-step response surface methodology. They showed that split-plot DSDs perform well in situations where there are a small number of significant factors.

4.8. OLS-GLS Equivalent Estimation Designs

Recall that the simplest split-plot model is given by $y = X\beta + \delta + \varepsilon$. Letting σ^2 and σ_δ^2 represent the sub-plot and whole plot error variances, and assuming that $\delta_k = N(0, \sigma_\delta^2)$ and $\varepsilon_{ij} = N(0, \sigma^2)$ are independent, the variance–covariance matrix of $\delta + \varepsilon$ is $\Sigma = \sigma^2 I + \sigma_\delta^2 J$, where J is a block diagonal matrix of 1_{bi} x $1'_{bi \, x \, 1}$, I is a block diagonal identity matrix, and b_i is the number of sub-plots within the ith whole-plot. Vining, Kowalski, and Montgomery[9] established the OLS and GLS estimates of β and the variance-covariance matrices for each of the estimates:

$$\widehat{\beta}_{OLS} = (X'X)^{-1}X'y \tag{6}$$

$$Var(\widehat{\beta}_{OLS}) = (X'X)^{-1}X'\Sigma X(X'X)^{-1} \tag{7}$$

$$\widehat{\beta}_{GLS} = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}y \tag{8}$$

$$Var(\widehat{\beta}_{GLS}) = (X'\Sigma^{-1}X)^{-1} \tag{9}$$

The OLS estimates is not the best linear unbiased estimator of β. The GLS estimate is the best linear unbiased estimator if and only if σ^2 and σ_δ^2 are known and if there is no OLS-GLS

equivalence. The estimates are equal if and only if a nonsingular matrix F exists such that $\boldsymbol{\Sigma X} = \boldsymbol{XF}$. When the OLS and GLS coefficient estimates are equal, the variance-covariance matrix for both coefficient estimates is equal.

Vining, Kowalski, and Montgomery[9] derived the necessary conditions to achieve OLS-GLS equivalent estimates for the regression coefficients and proposed two strategies conducive to achieving this condition. One strategy is to arrange each whole-plot with identical sub-plot designs, which typically result in large designs. The other strategy is to use a second-order orthogonal design with an identical number of sub-plot runs, which requires augmenting the design with center runs. Vining, Kowalski, and Montgomery[9] constructed OLS-GLS equivalent split-plot CCDs and BBDs, recommended options for obtaining balanced designs, and focused on estimating pure-error.

Bisgaard and Steinberg[97] and Bisgaard[7] used the equivalence between OLS-GLS for their first-order and first-order with interactions models. Bisgaard[7] achieved OLS-GLS equivalence in partial confounding designs even though not all sub-plots had the same experiment design. Conversely, Draper and John[98] and Trinca and Gilmour[85] recommended REML as alternative on how to estimate the variance components. REML is applicable to every possible split-plot design and provides good approximations for a good range of variance components; however, the variance component estimates depend on the specified model.

Second-order OLS-GLS equivalent estimation split-plot designs have received significant attention since 2005 since they have good features. Their construction is independent of a priori knowledge of the variance components. They can be analyzed using GLS algorithms, which are available in most commercial software packages. They are easy to generate. They provide pure-error estimates of the variance components that are independent of the model, which can be used for lack-of-fit tests, but that require an increased number of runs to make possible the estimation. Pure-error estimates are important in the early stages of experimentation. However, many practitioners object to replicating the center points under the precept that center points contribute very little towards building a model. Exact tests can be derived for at least some of the coefficients.

Parker, Kowalski, and Vining[10] provided a catalog of non-crossed OLS-GLS designs for equivalent estimation. Parker, Kowalski, and Vining[11] proposed methods for constructing balanced OLS-GLS equivalent estimation minimum whole-plot designs based on traditional response surface designs. The minimum whole-plot method is intended to reduce the number of whole-plots to the minimum number required to fit a second-order model by redistributing the runs that Vining, Kowalski, and Montgomery[9] allocated to overall center points to the whole-plot factorial points. Parker, Kowalski, and Vining[11] also provided a catalog of balanced and unbalanced CCDs and BBDs. Vining and Kowalski[99] established the appropriate error terms for testing pure sub-plot effects and effective whole-plot effects. Sub-plot residuals are though as individual data values predicted by the sub-plot model and adjusted by the whole-plot mean. Vining[14] wrapped it all together.

The designs by Vining, Kowalski, and Montgomery[9] have the most exact tests relative to minimum whole-plot designs and have unrestricted axial values for both whole factors (β) and sub-plot factors (α) axial points. They preserve the OLS-GLS equivalence with model reduction and in designs with whole-plots that only have center runs, but do not perform well relative to the D-optimality criteria due to the overabundance of runs at the center of the whole-plots. For minimum whole-plot designs, the OLS-GLS equivalence depends on α and β, and it is not preserved with model reduction or for designs that have whole-plots in which all runs are center runs.

Goos[100] provided an overview of blocked and split-plot designs and compared optimal designs and OLG-GLS equivalent estimation designs via estimation-based and prediction-based criteria. Goos[100] illustrated orthogonally blocked D-optimal designs and D-optimal split-plot designs for equivalent estimation. Macharia and Goos[101] provided D-Optimal and D-efficient equivalent estimation second-order split-plot designs. Jones and Goos[102] provided an algorithm for finding D-efficient equivalent estimation second-order split-plot designs that outperformed those of Macharia and Goos[101] and for cases in which the Macharia and Goos[101] algorithm couldn't find equivalent-estimation designs. Nguyen and Pham[103] described an algorithm that produced D-efficient equivalent estimation split-plot designs by interchanging the sub-plot factor levels within each whole-plot. They evaluated the design using the cases reported by Macharia and Goos[101] and Jones and Goos[102].

Yuan[103] constructed balanced and unbalanced OLS-GLS equivalent-estimation second-order split-split-plot CCDs. Aggarwal et. al.[104] constructed balanced and unbalanced split-plot CCDs and split-plot BBDs involving two quantitative sub-plot factors and one qualitative whole-plot factor based on the designs by Parker[10,11,12].

4.9. Blocked Split-Plot Designs

Sometimes, practitioners cannot complete an experiment under homogenous settings, and the variability associated with those settings permeates through the response variables and inflates the experimental error. This could be a problem since a precise comparison between and within treatments to detect the effects of the factors of interest requires homogeneous experimental units—a key concept introduced by Fisher[106]. Blocking is a form of local control of error. In a blocked design, the variability of the experimental units is less than the variability of the experimental units before they were grouped into blocks. A blocked design is complete if each block contains all of the treatments. Similarly, a blocked design is balanced if each block, which represents a level of the block factor, has an equal number of experimental units. Box and Hunter[48] and Khuri[107] exposed the conditions for orthogonal blocking in second-order designs. Khuri[107] generalized those conditions. Khuri[108] studied response surface models with fixed and random block effects and established that blocking increases the prediction variance. However, if the design blocks orthogonally, the blocks do not influence the estimation of the model coefficients and the least squares estimators of the regression variables are the same as without blocking. While some recent work by Tsai[109] has addressed blocking in first-order split-plot designs, only Wang, Kowalski, and Vining[110], Jensen and Kowalski[111], and Verma *et.al.*[112] incorporated blocking into second-order split-plot designs.

Wang, Kowalski, and Vining[110] constructed OLS-GLS equivalent central composite blocked split-plot designs and OLS-GLS equivalent Box-Behnken blocked split-plot designs from the second-order equivalent estimation designs proposed by Vining, Kowalski, and Montgomery[9] as well as from the minimum whole-plot designs proposed by Parker, Kowalski, and Vining[12]. While the designs by Wang, Kowalski, and Vining[110] have many appealing features and properties, the interaction between whole-plot factors is confounded with the block effect in cases when there

are only two whole-plot factors. Wang, Kowalski, and Vining[110] extended the second-order orthogonal blocking conditions to second-order split-plot designs.

Jensen and Kowalski[111] used a restricted CCD to fit a second-order model in a split-plot experiment involving two whole-plot factors and one sub-plot factor in the presence of blocking at sub-plot level. The design satisfied the conditions for OLS-GLS equivalent estimation. The experiment presented unique challenges for estimating the error terms and for checking the model assumptions. The parameters estimates were obtained using GLS although they could have been obtained using the simpler OLS estimation.

Verma *et.al.*[112] constructed balanced second-order blocked split-plot designs using designs by Dey[113] and unbalanced second-order blocked split-plot designs using designs by Zhang *et. al.*[114]. Dey[113] provided 3^k designs considering second-order orthogonal blocking. Zhang *et. al.* [114] provided small BBDs, but did not consider second-order orthogonal blocking. Verma *et. al.*[112] used a block size of two. The second-order split-plot designs derived from Dey[113] satisfied the second-order orthogonal blocking conditions, but the second-order split-plot designs derived from Zhang *et. al.*[114] did not. The algorithm to construct a second-order orthogonally blocked designs consisted of allocating sub-plots to whole-plots and whole-plots to blocks, sorting on certain factors, replicating whole-plots to achieve block balance, and then adding center runs to the whole-plots to obtain a second-order design that blocks orthogonally.

Baniani, Nargesi, Moghadam, and Wulff[115] examined the corrosion of medium carbon steel in an experiment with three factors in a split-block split-plot arrangement. In addition to showing the test of significance for this arrangement, the study made recommendations for fitting a second-order model.

Goos and Gilmour[88] showed how to carry out lack-of-fit tests for blocked, split-plot, or multi-stratum experiments and generalized the approach suggested by Vining, Kowalski, and Montgomery[9] and the tests proposed by Khuri[107] by exploiting replicates other than center point replicates. Arnouts and Goos[116] discussed the analysis of an experiment that involved the adhesion between steel tire cords and rubber, an ordinal response, and a random effect block factor in a split-plot structure.

4.10. Crossed and Cartesian Product Designs

Response surface split-plot designs began to receive significant attention for use in industrial experiments at the turn of the last century. Letsinger, Myers, and Lentner[6] investigated the effect that two hard-to-change factors (temperature 1 and pressure 1) and three easy to-change factors (humidity, temperature 2, and pressure 2) had on a (proprietary) response variable. A second-order model was expected to explain the relationships between the factors and the response. Letsinger, Myers, and Lentner[6] constructed both crossed response surface bi-randomization designs (BRD), which contain identical sub-plots in each whole-plot, and non-crossed response surface BRD, which may contain a different number of sub-plots in each whole-plot, and estimated the model regression coefficients using ordinary least squares (OLS), generalized least squares (GLS), iterated reweighted least squares (IRLS), and restricted maximum likelihood (REML). REML outperformed the other estimation techniques and OLS was appropriate only when the whole-plots were balanced. Letsinger, Myers, and Lentner[6] proved that OLS and GLS are equivalent if the sub-plot had the same experiment designs, but did not prove the equivalence with other conditions.

Vining[14] explained that for the cases reviewed by Letsinger, Myers, and Lentner[6], REML outperformed the other estimation techniques because the response surface designs were unbalanced. Because the designs were unbalanced, the OLS and GLS estimates were not equivalent; consequently, all the techniques for estimating the model coefficients are better estimators than OLS. Particularly, GLS is best-unbiased linear estimator if the whole-plot and sub-plot variances are known.

Cortes *et. al.*[117] provided an approach for constructing a response surface split-plot design referred to as *response surface Cartesian product split-plot design*. This type of design is constructed by crossing specific arrangements of whole-plot factors and sub-plot factors derived from CCDs, Box-Behnken designs, and definitive screening designs to generate response surface split-plot designs that are consistent with the traditional philosophy of response surface methodology. Response surface Cartesian product split-plot designs are economical, have a low prediction variance of the regression coefficients, and have low aliasing between model terms. In some cases, they can overcome some of the difficulties presented by other types of designs. Based

on an assessment using well accepted design evaluation criterion, response surface Cartesian product split-plot designs perform as well as historical designs that have been considered standards up to this point.

5. Response Surface Design Evaluation Criteria

The selection of an appropriate experiment design is often affected by factors such as the objective of the experiment, the homogeneity of the experimental units, the resources available to carry out the experiment, the complexity of the model to be fitted, and the capability to estimate internal error. Practitioners can select the most appropriate design by comparing different options over a wide range of characteristics.

5.1. General Design Evaluation Criteria

Box and Wilson[5] identified some characteristics of good response surface designs. Box and Hunter[48], Box and Draper[118], Box[119], and Box and Draper[120] further refined and expanded those characteristics, which include:

- distribute the information throughout the experimental region;
- provide a good fit of the model to the data;
- detect lack-of-fit;
- allow transformations;
- permit the experiment to be carried out in blocks;
- allow for the sequential assembly of higher-order designs;
- provide an estimation of internal error;
- be robust to outliers and the gross violation of normal theory assumptions;
- require a small number of experimental runs;
- provide data patterns that allow visual appreciation;
- ensure simple calculations;
- be robust to errors in control of factor levels;
- require a practical number of factor levels;
- check the homogeneous variance assumption;

Clearly, there are trade-offs in selecting a response surface design with good characteristics. Often, the experimental situation dictates the relative importance of those characteristics. While it is uncommon to find a design that simultaneously has all the characteristics listed above, a good design does not need to have them all. Most of the sources coincide in that a desirable property of response surface designs is a low and reasonably stable prediction variance over the design space (the scaled prediction variance measures the precision of the estimated response over the design space). The estimates are a function of the design, the model, and the location of the prediction in the design space. Park *et.al.*[121] discussed the prediction variance properties of second-order designs for cuboidal regions.

Box and Hunter[48] noted that criteria based only on the variances of the model terms was insufficient for the selection of a response surface design. Box and Draper[120] made clear the inherent danger of relying on only a single criterion and recommend choosing a design that balances many characteristics. Myers *et. al.*[77] pointed out that the importance of design robustness is underscored by forcing the use of a single criterion. Box and Draper[122] and Anderson-Cook, Borror, and Montgomery[75] are also useful references on the desired characteristics of response surface designs.

Myers, Montgomery, and Anderson-Cook[15] adapted the general guidelines to response surface split-plot designs. Like for response surface designs, a good response surface split-plot design should balance some of the following characteristics:
- provide a good fit of the model to the data;
- allow a precise estimation of the model coefficients;
- provide a good prediction over the experimental region
- provide an estimation of both whole-plot variance and sub-plot variance;
- detect lack-of-fit;
- check the homogeneous variance assumption at the whole-plot and sub-plot levels;
- consider the cost in setting the whole-plot and sub-plot factors;
- ensure the simplicity of the design;
- ensure simple calculations;
- be robust to errors in control of factor levels;
- be robust to outliers.

Lu, Anderson-Cook, and Robinson[123] used a multi-criteria pareto frontier approach to optimize the selection of a response surface design. Liang, Anderson-Cook, and Robinson[124] adapted the fraction of design space plots proposed by Zahran, Anderson-Cook, and Myers[125] to split-plot designs:

$$\frac{N}{\sigma_\delta^2 + \sigma^2} Var[\hat{y}(\mathbf{x})] = N\mathbf{x}'(\mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{x} \qquad (10)$$

5.2. Optimality Criteria

Optimality criterion provides a measure of how good a design is relative to a given objective function for a model. The criterion can be classified as information-based, distance-based, or compound. While those designs are optimal according to a single criterion for a specified statistical model, they could be sub-optimal according to another criterion. The designs are model dependent and may require a model that the user may not have. The efficiency of these designs depends on the number of factors, the number of points, and the maximum standard error for prediction over the design space. Typically, the best design for an application is the design with the highest optimality efficiency. The designs have designations corresponding to letters of the alphabet, such as D-, G-, I-, A-, V-, and E-optimality, to name a few. The most popular are the D-, G-, and I-optimal designs. D-optimal designs are good for screening while G- and I-optimal designs are good for characterization and optimization based on the variance properties. Many practitioners, undoubtedly, will eventually use some form of computer-generated optimal split-plot design where they would have the option to select the optimality criterion required by the objective of the experiment.

D-criterion attempts to minimize the variance of the regression coefficients $|(\mathbf{X}'\mathbf{X})^{-1}|$. G-criterion attempts to minimize the maximum scaled prediction variance over the design region R. I-criterion (also called Q- or IV-criterion) attempts to minimize the average scaled prediction variance by dividing $v(\mathbf{x})$ by the volume of R. The criteria for both a completely randomized design and a split-plot design are illustrated in Figure 2 where X_1 and X_2 represent the **X** matrices for each design, p represents the number of model parameters, and $Q(\xi)$ represents the scaled prediction variance averaged over the design region.

High D-efficiency is an indication of a good estimation of the model coefficients in terms of generalized variance. High G-efficiency is an indication of good prediction capability in terms of minimizing the maximum scaled prediction variance in the region of interest. High I-efficiency is an indication of good prediction capability in terms of the minimum average scaled prediction variance in the region of interest.

5.3. Cost

The cost of executing a completely randomized design is usually proportional to the overall number of runs because, typically, the cost of every treatment is essentially the same. This assumption does not hold in a split-plot experiment because often a split-plot experiment involves some factors that are costlier-to-change than others. Factors that are costlier-to-change are typically assigned to the whole-plot; thus, the cost driver for the experiment is the number of whole-plots. Because replication is needed to obtain an estimate of the whole-plot variance, practitioners tend to correlate this increase in overall sample size with an increase in cost. Therefore, it makes sense to judge the cost of a split-plot design by both the number of whole-plots and the number of runs within a whole-plot rather than by the number of total runs alone.

Bisgaard[7] used cost as part of a multiple criteria to compare the value of the information from the split-plot design against the cost of its runs. Parker, Anderson-Cook, Robinson, and Liang[126] demonstrated an approach that incorporates a cost function for evaluating the performance of competing second-order split-plot designs, and argued that the number of whole-plots is as important or more than the total number of runs. Additionally, the cost of blocking a split-plot experiment needs to factor in the blocking structure, the number of whole-plots, and the types of whole-plots.

6. Summary

Fisher[1] embedded the principles of replication, randomization, and local control of error in the fabric of experiment design and introduced the split-plot experiment for agronomic research. Box and Wilson[5] pioneered the application of design of experiments to industrial experiments and jump-started the development of response surface methodology. While response surface

methodology has experienced a significant growth since Box and Wilson[5], the growth of the design and analysis of second-order split-plot experiments, with and without blocking, has not received as much research attention. There is a vast body of literature related to response surface methodology, blocking, restricted randomization, design evaluation criteria, and first-order split-plot designs; however, literature on response surface split-plot design, particularly with blocking, is limited to only a few papers.

This literature research validates the need for improving industrial response surface split-plot design alternatives, without and with blocking. There is a need for improved approaches for constructing response surface split-plot designs (with and without blocking), especially for scenarios where replication must be minimized at the whole-plot level. Similarly, there is a need to refine the guidance and criteria for selecting better response surface split-plot designs. We encourage continued research in these areas.

7. References

1. Fisher, R. A. *The design of experiments*. Oliver and Boyd: Edinburgh, 1935.

2. Fisher, R. A. *Statistical methods for research workers*. Oliver and Boyd: Edinburgh, 1925.

3. Yates, F. The principles of orthogonality and confounding in replicated experiments. *Journal of Agricultural Science* 1933; **23**: 108-145.

4. Yates, F. Complex experiments. *Supplement to the Journal of the Royal Statistical Society* 1935; **2**: 181-247.

5. Box, G. E. P., and Wilson, K. G. On the experimental attainment of optimum conditions (with discussion). *Journal of the Royal Statistical Society B* 1951; **13**: 1-45.

6. Letsinger, J. D., Myers, R. H., and Lentner, M. Response surface methods for bi-randomization structures. *Journal of Quality Technology* 1996; **28**: 381-397.

7. Bisgaard, S. The design and analysis of 2^{k-p} x 2^{q-r} split-plot experiments. *Journal of Quality Technology* 2000; **32**: 39-56.

8. Goos, P. *The optimal design of blocked and split-plot experiments*. Springer: New York, NY, 2002.

9. Vining, G. G., Kowalski, S. M., and Montgomery, D. C. (2005). Response surface designs within a split-plot structure. *Journal of Quality Technology* 2005; **37**: 115-129.

10. Parker, P. A., Kowalski, S. M., and Vining, G. G. Classes of split-plot response surface designs for equivalent estimation. *Quality and Reliability Engineering International* 2006; **22**: 291-305.

11. Parker, P. A., Kowalski, S. M., and Vining, G. G. Construction of balanced equivalent estimation second-order split-plot designs. *Technometrics* 2007; **49**: 56-65.

12. Parker, P. A., Kowalski, S. M., and Vining, G. G. Unbalanced and minimal point equivalent estimation second-order split-plot designs. *Journal of Quality Technology* 2007; **39**: 376-388.

13. Jones, B., and Nachtsheim, C. J. Split-plot designs: What, Why, How. *Journal of Quality Technology* 2009; **41**: 340-361.

14. Vining, G. Split-plot response surface designs. In *Design and analysis of experiments, Vol. 3: Special designs and applications*. K. Hinkelmann (ed.). Wiley & Sons: Hoboken, NJ, 2012.

15. Myers, R. H., Montgomery, D. C., and Anderson-Cook, C. M. *Response surface methodology*. (3rd ed.). Wiley & Sons: New York, NY, 2016.

16. Hinkelmann, K., and Kempthorne, O. *Design and analysis of experiments: Volume 1 – Introduction to experimental design*. Wiley & Sons: New York, NY, 1994.

17. Federer, W. T., and King, F. *Variations on split-plot and split block experiment designs*. Wiley & Sons: New York, NY, 2007.

18. Wu, C. F.F., and Hamada, M. *Experiments; Planning, analysis, and optimization*. (2nd ed.). Wiley & Sons: Hoboken, NJ, 2009.

19. Montgomery, D. C. *Design and analysis of experiments*. (9th ed.). Wiley & Sons: New York, NY, 2017.

20. Kowalski, S. M., Parker, P. A., and Vining, G. G. Tutorial: Industrial split-plot experiments. *Quality Engineering* 2007; **19**: 1-15.

21. Bisgaard, S., and de Pinho, A. L. S. The error structure of split-plot experiments. *Quality Engineering* 2004; **16**: 671-675.

22. Searle, S. R., Casella, G., and McCulloch, C. E. *Variance components*. Wiley & Sons: New York, NY, 2006.

23. Huang, P., Chen, D., and Voelkel, J. Minimum aberration two-level split-plot designs. *Technometrics* 1998; **40**: 314-326.

24. Bingham, D. R., and Sitter, R. R. Some theoretical results for fractional factorial split-plot designs. *Annals of Statistics* 1999; **27**: 1240-1255.

25. Bingham, D. R., and Sitter, R. R. Minimum aberration two-level fractional-factorial split-plot designs. *Technometrics* 1999;**41**: 62-70.

26. Bingham, D. R., and Sitter, R. R. Design issues in fractional factorials split-plot experiments. *Journal of Quality Technology* 2001;**33**: 2-15.

27. Loeppky, J. L., and Sitter, R. R. Analyzing unreplicated blocked or split-plot fractional factorial designs. *Journal of Quality Technology* 2002; **34**: 229-243.

28. Addelman, S. Some two-level factorial plans with split-plot confounding. *Technometrics* 1964; **6**: 253-258.

29. Kulahci, M., and Bisgaard, S. The use of Placket Burman Designs to construct split-plot designs. *Technometrics* 2005; **47**: 495-501.

30. Plackett, R. L, and Burman, J. P. The design of optimum multifactorial experiments. *Biometrika* 1946; **33**: 305-325, 328-332.

31. Tyssedal, J., Kulahci, M. Analysis of split-plot designs with mirror image pairs as sub-plots. *Quality and Reliability Engineering International* 2005; **21**: 539-551.

32. Tyssedal, J., Kulahci, M., and Bisgaard, S. Split-plot designs with mirror image pairs as sub-plots. *Journal of Statistical Planning and Inference* 2011; **41**: 3686-3696.

33. Kulahci, M, and Tyssedal, J. Split-plot designs for multistage experimentation. *Journal of Applied Statistics* 2017; **44**: 493-510.

34. Hader, R. J. An improper method of randomization in experimental design. *The American Statistician* 1973; **27**: 82-84.

35. Wooding, W. M. The split-plot design. *Journal of Quality Technology* 1973; **5**: 16-33.

36. Box, G. E. P. Split plot experiment. Quality Engineering 1996; **8**: 515-520.

37. Simpson, J. R., Kowalski, S. M., and Landman, D. Experimentation with randomization restrictions: Targeting practical implementation. *Quality and Reliability Engineering International* 2004; **20**: 481-495.

38. Daniel, C. *Applications of statistics to industrial experimentation*. Wiley & Sons: New York, NY, 1976.

39. Ganju, J., and Lucas, J. M. Bias in test statistics when restrictions in randomization are caused by factors. *Communications in Statistics, Theory and Methods* 1997: **26**: 47-63.

40. Ganju, J., and Lucas, J. M. Detecting randomization restrictions caused by factors. *Journal of Statistical Planning and Inference* 1999: **81**: 129-140.

41. Ju, H. L., and Lucas, J. M. Lk factorial experiments with hard-to-change and easy-to-change factors. *Journal of Quality Technology* 2002; **34**: 411-421.

42. Webb, D. F., Lucas, J. M., and Borkowski, J. J. (2004). Factorial experiments when factor levels are not necessarily reset. *Journal of Quality Technology* 2004; **36**: 1-11.

43. Taguchi, G. *System of experimental design, Vols. 1 and 2*. Kraus International: White Plains, NY, 1987.

44. Box, G. E. P., and Jones, S. Split-plot designs for robust product and process experimentation. *Journal of Applied Statistics* 1992; **19**: 3-26.

45. Bisgaard, S., and Sutherland, M. Split-plot experiments: Taguchi's Ina tile experiment reanalyzed. *Quality Engineering* 2003; **16**: 157-164.

46. Box, G. E. P., and Jones, S. Split-plots for robust product and process experimentation. *Quality Engineering* 2001:**13**: 127-134.

47. Box, G. E. P., and Behnken, D. W. Some new three-level designs for the study of quantitative variables. *Technometrics* 1960; **2**: 455-476.

48. Box, G. E. P., and Hunter, J. S. Multi-factor experimental designs for exploring response surfaces. *Annals of Mathematical Statistics* 1957; **28**: 195-241.

49. Hartley, H. O. Smallest composite designs for quadratic response surfaces. *Biometrics* 1959; **15**: 611-624.

50. Westlake, W. J. Composite designs based on irregular fractions of factorials. *Biometrics* 1965; **21**: 324-336.

51. Rechtschaffner, R. l. Saturated fractions of 2^n and 3^n factorial designs. *Technometrics* 1967; **9**: 569-575.

52. Doehlert, G. H. Uniform shell designs. *Journal of the Royal Statistical Society, Series C* 1970; **19**: 231-239.

53. Hoke, A. T. Economical second-order designs based on irregular fractions of the 3n factorial. *Technometrics* 1974; **16**: 375-384.

54. Pesotchinsky, L. L. D-optimum and quasi D-optimum second-order designs on a cube. *Biometrika* 1975; **62**: 335-340.

55. Lucas, J. M. Optimum composite designs. *Technometrics* 1974; **16**: 561-567.

56. Box, G. E. P., and Draper, N. R. On minimum-point second-order designs. *Technometrics* 1974; **16**: 613-616.

57. Roquemore, K. G. Hybrid designs for quadratic response surfaces. *Technometrics* 1976; **4**: 419-423.

58. Mitchell, T. J., & Bayne, C. K. D-optimal fractions of three level factorial designs. *Technometrics* 1978; **20**: 369-380.

59. Notz, W. Minimal point second-order designs. *Journal of Statistical Planning and Inference* 1982; **6**: 47-58.

60. Draper, N. R. Small composite designs. *Technometrics* 1985; **27**: 173-180.

61. Morris, M. A class of three-level experimental designs for response surface modeling. *Technometrics* 2000; **42**: 111-121

62. Oehlert, G. W., and Whitcomb, P. Small, efficient, equireplicated resolution V fractions of 2^k designs and their application to central composite designs. *46th Annual Fall Technical Conference* 2002, ASQ and ASA: Valley Forge, PA.

63. Gilmour, S. G. Irregular four-level response surface designs. *Journal of Applied Statistics* 2004; **31**: 1043-1048.

64. Kiefer, J. C. Optimum experimental design (with discussion). *Journal of the Royal Statistical Society B* 1959; **21**: 272-319.

65. Kiefer, J. C. Optimum designs in regression problems. II. *Annals of Mathematical Statistics* 1961; **32**: 298-325.

66. Keifer, J., and Wolfowitz, J. The equivalence of two extremum problems. *Canadian Journal of Mathematics* 1960; **12**: 363-366.

67. Jones, B., and Nachtsheim, C. J. A class of three level designs for definitive screening in the presence of second-order effects. *Journal of Quality Technology* 2011; **43**: 1-15.

68. Jones, B., and Nachtsheim, C. J. Definitive screening designs with added two-level categorical factors. *Journal of Quality Technology* 2013; **45**; 120–129.

69. Jones, B., and Nachtsheim, C. J. Blocking schemes for definitive screening designs. *Technometrics* 2015; **58**: 74-83.

70. Jones, B. 21st century screening experiments: What, why, and how? (With discussion). *Quality Engineering* 2016; **28**: 98-126

71. Errore, A., Jones, B., Li, W., and Nachtsheim, C. J. Benefits and fast construction of efficient fold over designs. *Technometrics* 2017; **59**: 48-57.

72. Errore, A., Jones, B., Li, W., and Nachtsheim, C. J. Using definitive screening designs to identify active first- and second-order factor effects. *Journal of Quality Technology* 2017; **49**: 244-264.

73. Jones B., and Nachtsheim, C. J. Effective design-based model selection for definitive screening designs, *Technometrics* 2017; **59**: 319-329.

74. Myers, R. H. Response surface methodology; current status and future directions. *Journal of Quality Technology* 1999; **31**: 30-44.

75. Anderson-Cook, C. M., Borror, C. M., and Montgomery, D. C. Response surface design evaluation and comparison (with discussion). *Journal of Statistical Planning and Inference* 2009; **139**: 629-674.

76. Neff, A. A., and Myers, R. H. Recent development in response surface methodology and its applications in industry. In *Statistical process monitoring and optimization*. S. H. Park and G. G. Vining (eds.). Marcel Dekker: New York, NY, 2000; 457-481.

77. Myers, R. H., Montgomery, D. C., Vining, G. G., Borror, C. M., and Kowalski, S. M. Response surface methodology: A retrospective and literature survey. *Journal of Quality Technology* 2004; **36**: 53-77.

78. Khuri, A. I., and Mukhopadhyay, S. Response surface methodology. *WIREs Computational Statistics* 2010; **2**: 128-149.

79. Khuri, A.I. Response surface methodology and its applications in agricultural and food sciences. *Biometrics & Biostatistics International Journal* 2017; **5**: 141-152.

80. Lucas, J. M., and Ju, H. L. Split-plotting and randomization in industrial experiments. *ASQC Quality Congress Transactions* 1992; 374-382.

81. Wang, L., Vining, G. G., and Kowalski, S. M. Two-strata rotatability in split-plot central composite designs. *Applied Stochastic Models in Business and Industry* 2010; **26**: 431-447.

82. English, T. G., Simpson, J. R., Landman, D., and Parker, P. A. An efficient split-plot approach for modeling non-linear aerodynamic effects. *Quality Engineering* 2012; **24**: 522-530.

83. Cheng, C-S. *Theory of factorial design; Single- and multi-stratum experiments*. CRC Press: Boca Raton, NY, 2014.

84. Gilmour, S. G., and Trinca, L. A. Some practical advice on polynomial regression analysis from blocked response surface designs. *Communications in Statistics: Theory and Methods* 2000; **29**: 2157–2180.

85. Trinca, L. A., and Gilmour, S. G. Multistratum response surface designs. *Technometrics* 2001; **43**: 25-33.

86. Gilmour, S. G. Response surface designs for experiments in bioprocessing. *Biometrics* 2006; **62**: 323–331.

87. Mylona, K., Macharia, H., and Goos P. Three level equivalent estimation split-plot designs based on subset and supplementary difference set designs. *University of Antwerp website*. http://www.ua.ac.be/tew, D/2011/1169/010 [14 December 2017]

88. Goos, P., Gilmour, S. G. Testing for lack-of-fit in blocked, split-plot, and other multi-stratum designs. *Journal of Quality Technology* 2017; **49**: 320-336

89. Gilmour, S. G., and Trinca, L. A. Optimum design of experiments for statistical inference (with discussion). *Applied Statistics* 2012; **61**: 345–401.

90. Trinca, L. A., and Gilmour, S. G. Improved split-plot and multi-stratum response surface designs. *Technometrics* 2015; **57**: 145–154.

91. Trinca, L. A., and Gilmour, S. G. Split-plot and multistratum designs for statistical inference. *Technometrics* 2017; **59**: 446-457.

92. Sartono, B., Goos, P., and Schoen, E. Constructing general orthogonal fractional factorial split-plot designs. *Technometrics* 2015; **57**: 488-502.

93. Goos, P., and Vandebroek, M. Optimal split-plot designs. *Journal of Quality Technology* 2001; **33**: 436-450.

94. Goos, P., and Vandebroek, M. D-optimal split-plot designs with given number and sizes of whole-plots. *Technometrics* 2003; **45**: 235-245.

95. Jones, B., and Goos, P. I-optimal versus D-optimal split-plot response surface designs. *University of Antwerp website*. http://www.ua.ac.be/tew, D/2012/1169/002 [14 December 2017]

96. Lin, C-Y., Yang, P. Response surface methodology using split-plot definitive screening designs. *Journal of Quality Technology* 2015; **47**: 351-362.

97. Bisgaard, S., and Steinberg, D. M. The design and analysis of 2^{k-p} x s prototype experiments. *Technometrics* 1997; **39**:52-62.

98. Draper, N. R., and John, J. A. Response surface designs where levels of some factors are difficult to change. *Australian and New Zealand Journal of Statistics* 1998; **40**: 487-495.

99. Vining, G. G., and Kowalski, S. M. Exact inference for response surface designs within a split-plot structure. *Journal of Quality Technology* 2008; **40**: 394-406.

100. Goos P. Optimal versus orthogonal and equivalent-estimation design of blocked and split-plot experiments. *Statistica Nerlandica* 2006; **60**: 361-378.

101. Macharia, H., and Goos, P. D-Optimal and D-efficient equivalent estimation second-order split-plot designs. *Journal of Quality Technology* 2010; **42**: 358-372.

102. Jones, B., and Goos, P. An algorithm for finding D-efficient equivalent-estimation second-order split-plot designs. *Journal of Quality Technology* 2012; **44**: 363-374.

103. Nguyen, N-K., Pham, T-D. Searching for D-efficient equivalent estimation second-order split-plot designs. *Journal of Quality Technology* 2015; **47**: 54-65.

104. Yuan, F., Perry, M. B., Conerly, M. D., Adams, M. B., Lee, J., Raja, U. Construction of equivalent estimation second-order split-split-plot designs. *Doctoral dissertation*. University of Alabama Website http://acumen.lib.ua.edu/ [15 December 2017]

105. Aggarwal, M., Chowdhury, S. R., Bansal, A., and Verma A. Balanced and unbalanced response surface designs involving qualitative factors under split-plot. *Journal of the Indian Society of Agricultural Statistics* 2014; **68**: 25-38.

106. Fisher, R. A. The arrangement of field experiments. *Journal of the Ministry of Agriculture of Great Britain* 1926; **33**: 503-513.

107. Khuri, A. I. Response surface models with random block effects. *Technometrics* 1992; **34**: 26-37.

108. Khuri, A. I. Current modeling and design issues in response surface methodology: GLMs and models with block effects. In *Handbook of Statistics*, Vol 22, Chapter 6. R. Khattree and C. R. Rao (eds). Elsevier Science: Amsterdam, 2003.

109. Tsai, P-W. A study of two types of split-plot designs. *Journal of Quality Technology* 2016; **48**: 44-53.

110. Wang, L., Kowalski, S. M., and Vining, G. G. Orthogonal blocking of response surface split-plot designs. *Journal of Applied Statistics* 2009; **36**: 303-321.

111. Jensen, W. A., and Kowalski, S. M. Response surfaces, blocking, and split-plots: An industrial experiment case study. *Quality Engineering* 2012; **24**: 531-542.

112. Verma, A., Aggarwal, M. L., Chowdhury, S. R., Singh, P. Response surface designs under split-plot structure. *Doctoral dissertation*. Shodhganga, Indian ETD Repository Website. http://hdl.handle.net/10603/28352 [7 April 2015]

113. Dey, A. Orthogonally blocked three-level second-order designs. *Journal of Statistical Planning and Inference* 2009; **139**: 3698-3705.

114. Zhang, T., Yang, J., and Lin, D. K. J. Small Box-Behnken designs. *Statistics and Probability Letters* 2011; **81**: 1027-1033

115. Baniani, A. M., Nargesi, Z. R., Moghadam, A. M., and Wulff, S. S. Split-block split-plot experimental design to assess corrosion in medium carbon steel. *Quality Engineering* 2015; **27**: 212-229.

116. Arnouts, H., Goos, P. Analyzing ordinal data from a split-plot design in the presence of a random block effect. *Quality Engineering* 2017; **29**: 553-562.

117. Cortes, L. A., Simpson, J. R., Duff, W. S., Chong, K. P., Bradley, T. H., and Jathar, S. H. Second-order sub-array Cartesian product split-plot design. *Doctoral dissertation*. Colorado State University 2015. https://dspace.library.colostate.edu [19 December 2017]

118. Box, G. E. P., and Draper, N. R. A basis for the selection of a response surface design. *Journal of the American Statistical Association* 1959; **72**: 622-654.

119. Box, G. E. P. Response surfaces. In *The International Encyclopedia of the Social Sciences*. D. L. Sills (ed). MacMillan and Free Press: New York, NY, 1968.

120. Box, G. E. P., and Draper, N. R. Robust designs. *Biometrika* 1975; **62**: 347-352.

121. Park, Y. -J., Richardson, D. E., Montgomery, D. C., Ozol-Godfrey, A., Borror, C. M., Anderson-Cook, C. M. Prediction variance properties of second-order designs for cuboidal regions. *Journal of Quality Technology* 2005; **37**: 253-266.

122. Box, G. E. P., and Draper, N. R. *Response surfaces, mixtures, and ridge analysis* (2nd ed.). Wiley & Sons: New York, NY, 2007.

123. Lu, L., Anderson-Cook, C., and Robinson, T. J. Optimization of designed experiments based on multiple criteria utilizing a pareto frontier. *Technometrics* 2011; **54**: 353–365

124. Liang, L., Anderson-Cook, C. M., and Robinson, T. J. Fraction of design space plots for split-plot designs. *Quality and Reliability Engineering International* 2006; **22**: 275-289.

125. Zahran, A., Anderson-Cook, C. M., and Myers, R. H. Fraction of design space to assess the prediction capability of response surface designs. *Journal of Quality Technology* 2003; **35**: 377-386.

126. Parker, P. A., Anderson-Cook, C., Robinson, T. J., and Liang, Li. Robust split-plot designs. *Quality and Reliability Engineering International* 2007; **24**: 107-121.